

CIRM Genomics Knowledge Network

David Haussler
UC Santa Cruz Genomics Institute
February 22, 2021

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

CIRM Center For Excellence in Stem Cell Genomics

- Centerpiece of CIRM Knowledge Network is the Stem Cell Hub of the Center for Excellence in Stem Cell Genomics.
- The Center is led by Mike Snyder (Stanford), Joe Ecker (Salk), and Josh Stuart (UCSC)
- Focus: **Cardiovascular Disease**, **Cell Differentiation** in brain and pancreas, **Molecular Networks** in all types of stem cells
- Supports 15 CIRM laboratories in addition to Snyder, Ecker and Stuart
- All genomics data generated by supported projects are made available globally through the Stem Cell Hub no later than the time of publication.

The Stem Cell Hub
Center for Excellence in Stem Cell Genomics

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

CIRM
CALIFORNIA STEM CELL AGENCY

Stem Cell Hub: Data

https://cirm.ucsc.edu

84 TB of data in 180,170 files from 18 CIRM research labs

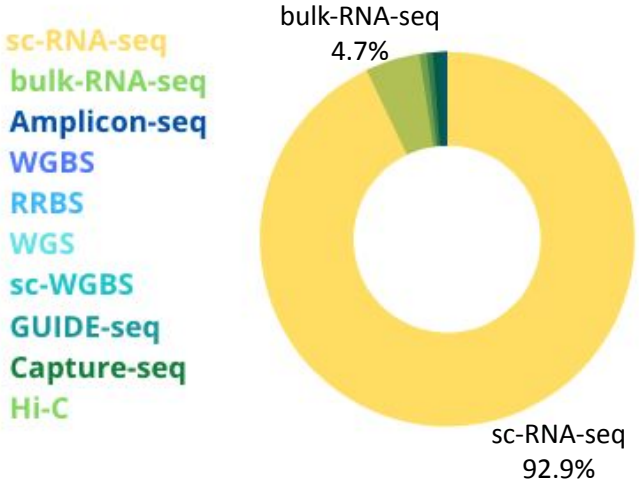
6.7 TB of that data is **currently publicly available;** the rest is pre-publication

The data are **Machine-Learning Ready***

*https://www.acd.od.nih.gov/documents/presentations/12132019AI_FinalReport.pdf

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

Stem Cell Hub: Assays & Organs



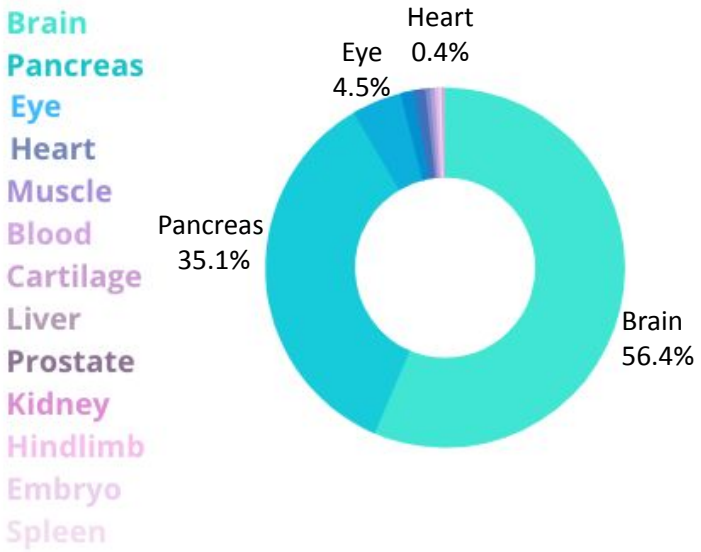
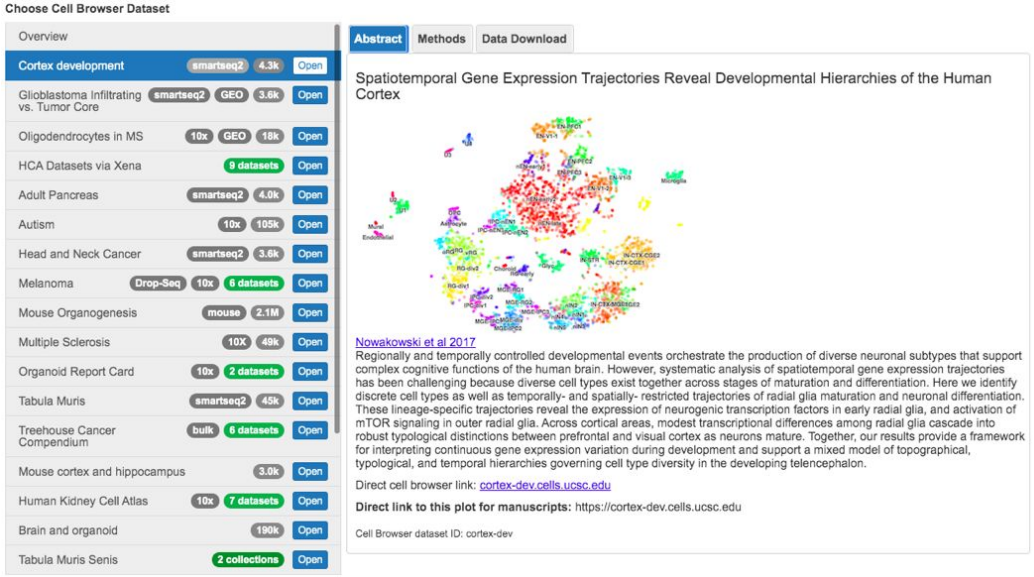


**chart is based on proportion of fastq files in each category*

Cell Browser

The Cell Browser is a software **tool using a 2D viewer** to represent single-cell RNA expression



https://cells.ucsc.edu/

Integrates CIRM data with global single-cell data, including HCA

Shows **expression data for individual cells**

Allows for a **visual comparison** of large datasets consisting of many cells

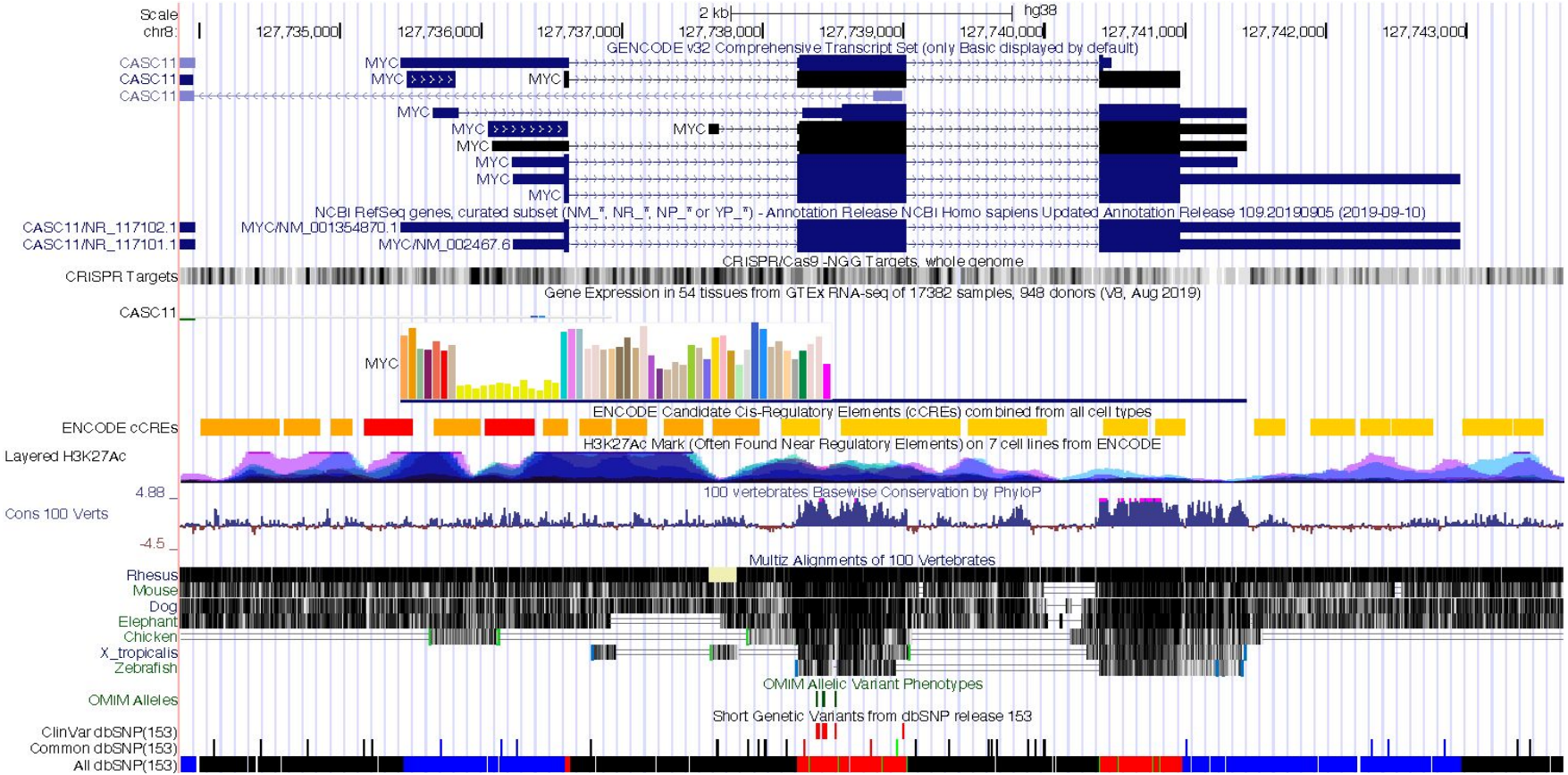
Includes overlays of **metadata, marker gene levels, cell clustering and more**

Useful for comparing single-cell layout/batch correction methods

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

Coordinated with Human Cell Atlas (HCA)

Genome Browser



https://genome.ucsc.edu/

Data Biosphere

Scalable and interoperable computing resource for the genomics scientific community

Cloud-based infrastructure

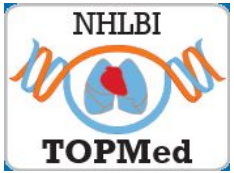
- Highly elastic; shared analysis and computing environment

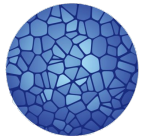
Data access and security

- Genomic and single cell datasets, phenotypes and metadata
- Securely housed, large datasets generated by NHGRI, NHLBI, NCI, CZI funded programs and HCA community, as well as other initiatives / agencies

Collaborative computing environment for datasets and analysis workflows

- Storage, scalable analytics, data visualization
- Security, training & outreach, with new models of data access
  - ...for both users with limited computational expertise and sophisticated data scientist users





HUMAN CELL ATLAS

NIH National Heart, Lung, and Blood Institute

BioData CATALYST

AnVIL

DATA BIOSPHERE

The next phase in genomics: a complete reference of human genome diversity

HUMAN PANGENOME

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

TELOMERE-TO-TELOMERE CONSORTIUM

NIH Genomics

June & July 2000
First draft of human genome sequence released

2005
International HapMap Project

2006
The Cancer Genome Atlas

1000 Genome Project

2008

2010
Draft Sequence of the Neandertal

2013
Global Alliance for Genomics and Health (GA4GH)

2015
Launch USA All of Us Biobank

2012
Electronic Medical Records and Genomics (eMERGE)

biobank UK
Applicants Accepted: UK Biobank

2014
The Human Heredity and Health in Africa (H3Africa)

Trans-Omics for Precision Medicine (TOPMed) Program

2016
Simons Diversity Study
GWAS Catalog
The Human Cell Atlas

2017
Population Architecture using Genomics and Epidemiology (PAGE)

Implementing Genomics in Practice (IGNITE)

2019
Launch of the Human Pangenome Reference Consortium

HUMAN PANGENOME

2020

NIH Genomics

June & July 2000
First draft of human genome sequence released

2004
CIRM created

2005
International HapMap Project

2006
The Cancer Genome Atlas

1000 Genome Project

2008

2010
Draft Sequence of the Neandertal

2012
Electronic Medical Records and Genomics (eMERGE)

biobank UK
Applicants Accepted: UK Biobank

2013
Global Alliance for Genomics and Health (GA4GH)

2014
The Human Heredity and Health in Africa (H3Africa)

Trans-Omics for Precision Medicine (TOPMed) Program

2015
Launch USA All of Us Biobank

2016
Simons Diversity Study
GWAS Catalog
The Human Cell Atlas

2017
Population Architecture using Genomics and Epidemiology (PAGE)

Implementing Genomics in Practice (IGNITE)

2019
Launch of the Human Pangenome Reference Consortium

HUMAN PANGENOME

2020

NIH Genomics

New initiative to launch
The development of a new pan-reference genome!

June & July 2000
First draft of human genome sequence released

2004
CIRM created

2005
International HapMap Project

2006
The Cancer Genome Atlas

1000 Genome Project

2008

2010
Draft Sequence of the Neandertal

2012
Electronic Medical Records and Genomics (eMERGE)

biobank UK
Applicants Accepted: UK Biobank

2013
Global Alliance for Genomics and Health (GA4GH)

2014
The Human Heredity and Health in Africa (H3Africa)

Trans-Omics for Precision Medicine (TOPMed) Program

2015
Launch USA All of Us Biobank

2016
Simons Diversity Study
GWAS Catalog
The Human Cell Atlas

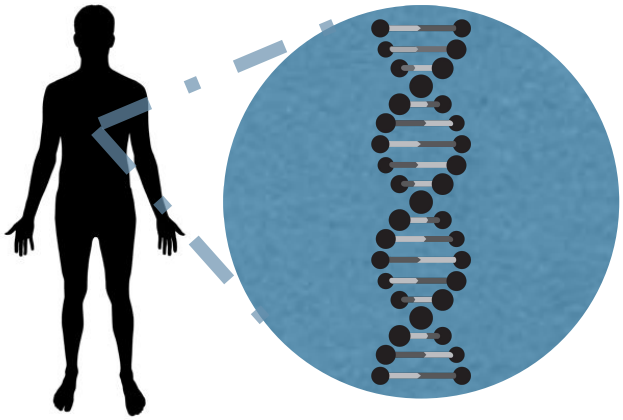
2017
Population Architecture using Genomics and Epidemiology (PAGE)
Implementing Genomics in Practice (IGNITE)

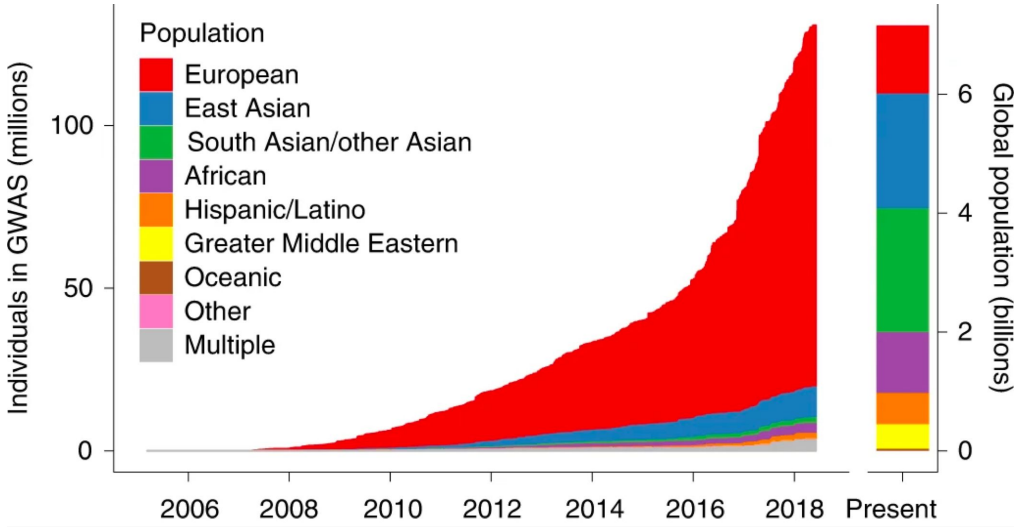
2019
Launch of the Human Pangenome Reference Consortium

HUMAN PANGENOME

2020

Human Reference Genome: Largely represents one individual





The Human Reference Genome:

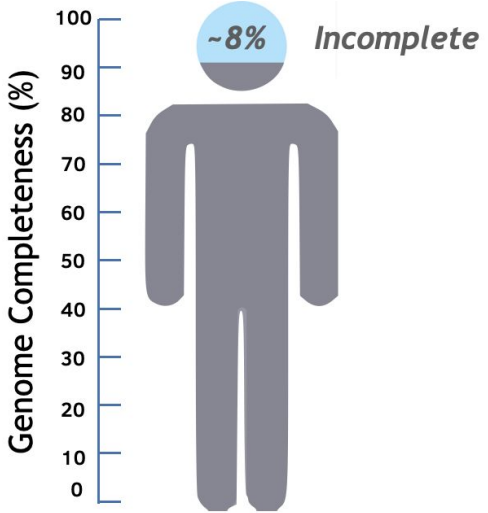
Does not adequately represent genetic diversity in the human population.

Martin et al 2019

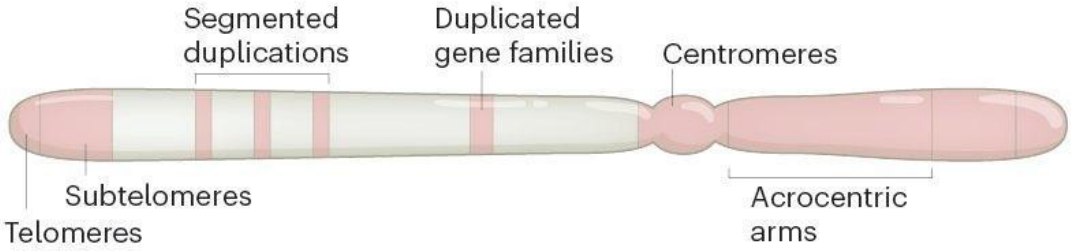
A Human Pangenome Reference Eliminates Disparities

- Unbiased representation of sequence diversity in the human population
- Comprehensive map of genome variation
- New reference data structure to nucleate and foster a new ecosystem of pangenome tools for clinical use

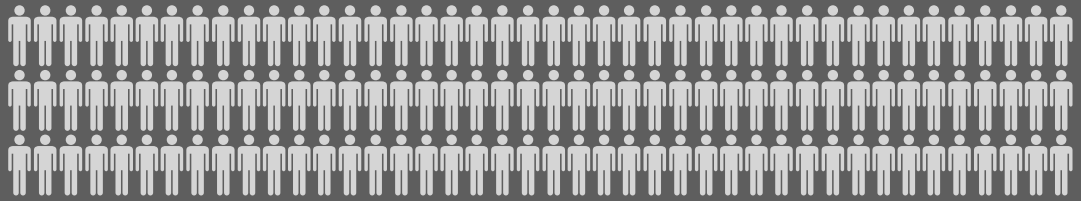
Need to generate and analyze *complete* human genomes



Human Genome Project (April 2003):
Focused Exclusively on Finishing 99% *Euchromatic* Regions.
Highly-Repetitive Heterochromatin Regions were not Included



Miga, Nature 2020

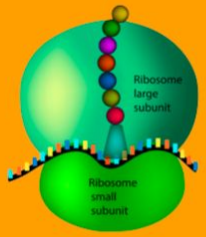


Technological Challenge:
Production of *Complete* Reference Genomes

GENOME INSTABILITY AND GENE FAMILIES

Large, Unresolved Intra and Inter Segmental Duplications

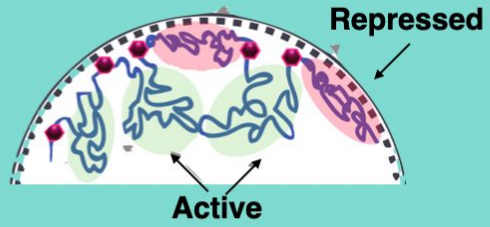
RIBOSOMES: TRANSLATION



Bringing together amino acids to form particular proteins

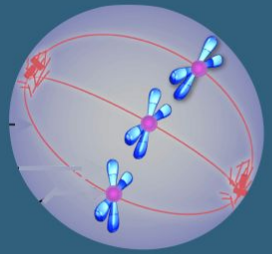
rDNA Arrays

GENOME SPATIAL ORGANIZATION

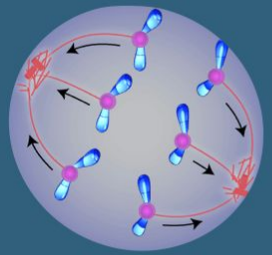


Typically Spatially distinct from active TAD compartments

CENTROMERE FUNCTION



Contribute to Chromosome Cohesion

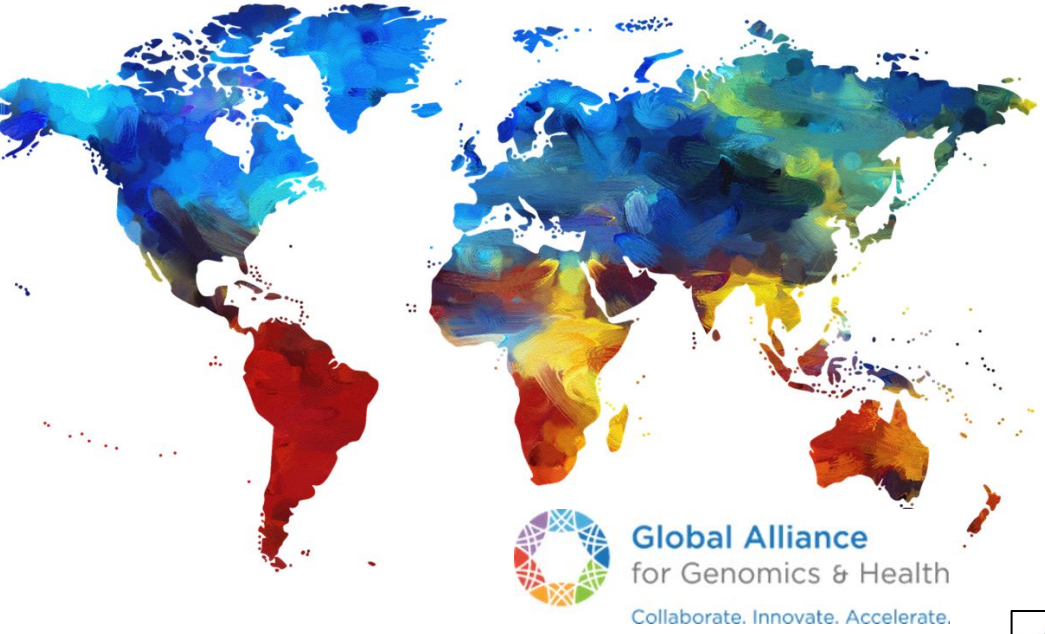


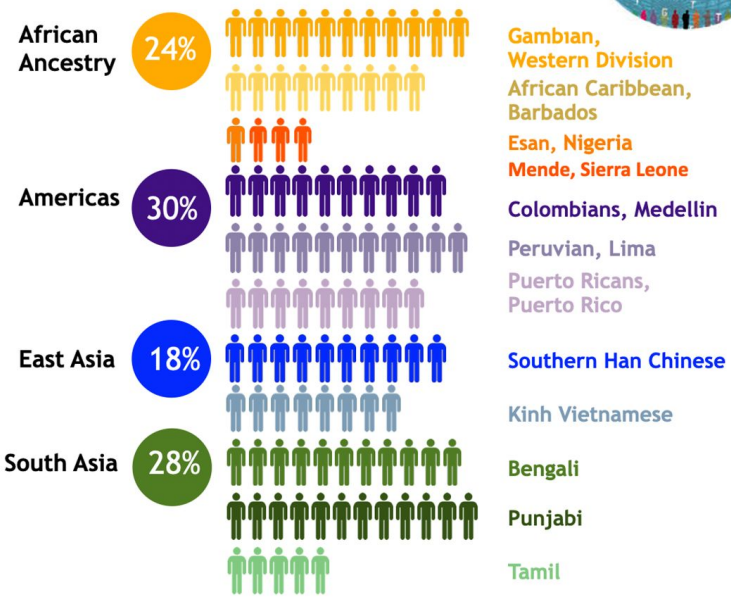
Regulate Centromere Function

NIH Pangenome Project Population

Representation and Sampling

Global Genomic Partnerships





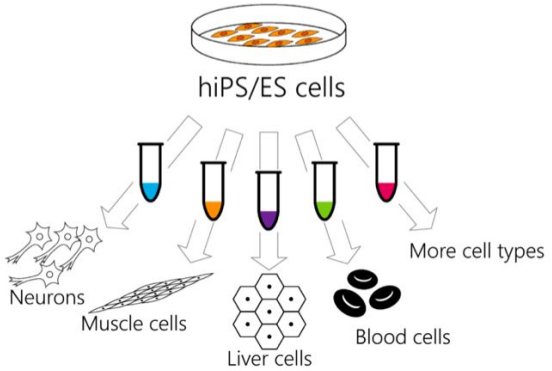
New partnerships (domestic & international) to reach a more “complete” human pangenome reference

- ✓ Cover genetic and geographic diversity
- ✓ Availability of low passage cell lines
- ✓ Availability of trios/parental data.

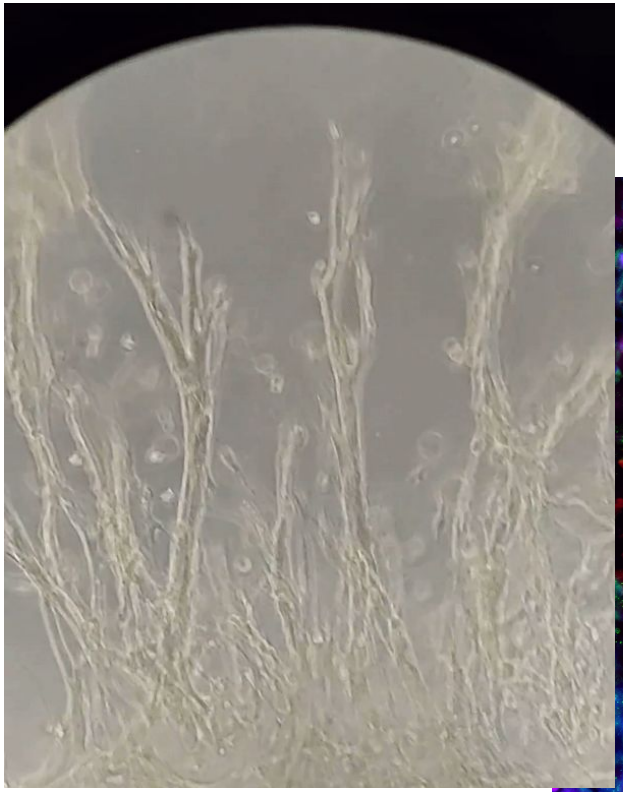
CIRM Stem Cell Opportunities

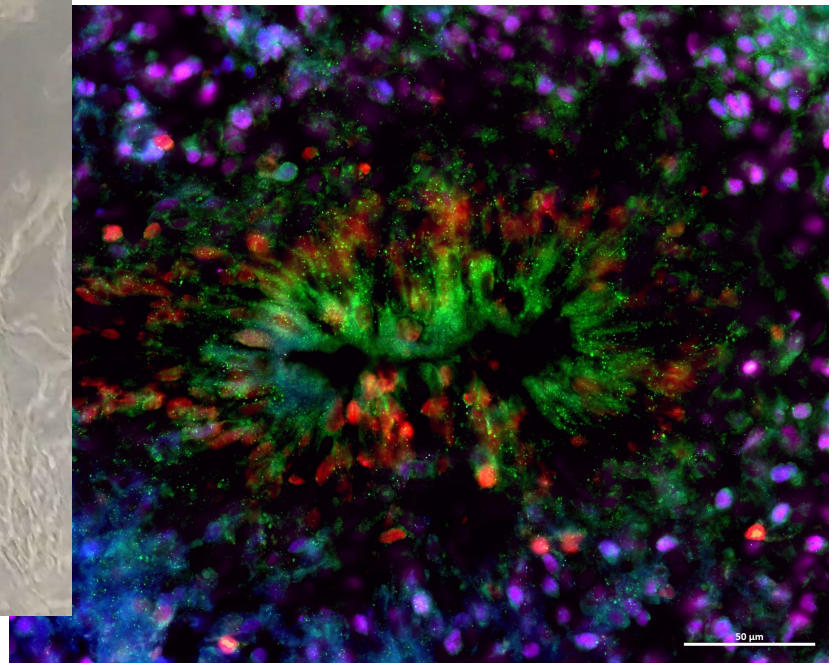
- A **reference iPSC** line will be needed for each of the **>350 diverse globally recognized reference genomes funded by NIH** with unrestricted use
- Possibility to establish and distribute **standard organoids** of various tissue types derived from reference lines, an "**organoid nursery**": **Organoid Hotel California**
- Opportunity to construct the first "**Human Pan-Epigenome**" by concentrating global research on a common reference lines and organoids, validate with primary tissue
- Opportunity for collaboration with many **NIH institutes making disease-specific iPSC libraries**
- Ethnically diverse reference iPSCs could serve as **matched normals for wide range of disease cohort studies**

Human Pangenome iPSC Diversity Panel:



Example: Cerebral Organoids



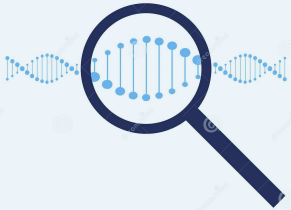


Green: HOPX
Red: SOX2
Lavender: CTIP2
Blue: SATB2

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

Ryan Hoffman, UCSC

The marriage of stem cell and genomics research is the right foundation for a more equitable regenerative medicine

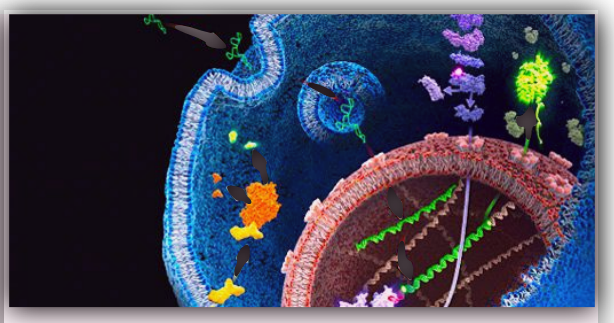


Health Care

Ability to study a *complete genome* for important clinical variants

Genome Diversity

Improve our understanding of human sequence variation and evolution



Epigenome

Expand epigenetic studies across different cell types for important clinical variants

Acknowledgements

UC Santa Cruz Genomics Institute

Jim Kent
Benedict Paten
Karen Miga
Ed Green
Mark Akeson
Adam Novack
Miten Jain
Hugh Olsen
Erik Garrison
Marina Hauknes
Jean Monlongs
Glenn Hickey
Adirna Fuller

Tony Tsung Yu Lu
Xian Chang
Trevor Pesout
Ryan Lorig-Roach
Charles Markello
Melissa Meredith
Jonas Sibbesen
Kishwar Shafin
Jouni Siren
Beth Sheets
Jordan Eizenga
Julian Lucas
Brian Hannafious

Galt Barber
Jonathan Casper
Max Haeussler
Clay Fischer
Parisa Nejad
Matthew Speir
William Sullivan
Chris Villarreal
Kamron Mojabe
Pranav Muthuraman
Beagan Nguy
Tiana Pereira

The Genome Institute at Washington University

Ting Wang
Lucinda Fulton
Sarah Cody
Robert Fulton
Wen-Wei Liao
Nathan Stitziel
Haley Abel

Eddie Belter
Milinn Kremitizki
Derek Albracht
Chad Tomlinson
Allison Regier
Chris Markovic
Tina Lindsay

EMBL-EBI

Paul Flicek
Susan Fairley
Daniel Zerbino

NIH National Human Genome Research Institute

Adam Felsenfeld
Mike Smith
Carolyn Hutter
Taylorlyn Stephan
Heidi Sofia

Adam Phillippy
Sergey Koren
Arang Rhie
Chirag Jain
Baergen Schultz

NIST

Justin Zook

Wellcome Sanger Institute

Kerstin Howe

Icahn School of Medicine at Mount Sinai

Eimear Kenny
Vimi Desai

CBG Max Planck Institute of Molecular Cell Biology and Genetics

Gene Myers

NCBI National Center for Biotechnology Information

Valerie Schneider
Terence Murphy
Paul Kitts
Chunlin Xiao
Francoise Thibaud-Nissent

Genome Sciences University of Washington

Evan Eichler
Katy Munson
Mitchell Vollger

Pete Audano
David Porubksy
Arvis Sulovari

Coriell Institute for Medical Research
Decoding the Genome

Alissa Resch
Brittney Martinez
Brittany Kerr
Ellen Kellly

Yale University

Ira Hall
Wen-Wei Liao
Shuangjia Lu

Harvard Medical School

Heng Li
Shilpa Garg
Haoyu Cheng
Xiaowen Feng

University of Cambridge

Richard Durbin

The Rockefeller University
Science for the benefit of humanity

Erich Jarvis
Lauren Shalmiyev
Olivier Fedrigo

Giulio Forment
Sadye Paezi

Company Partnerships

AnVIL
aws
illumina
PACBIO
HudsonAlpha Institute for Biotechnology
Google DeepVariant
ARIMA Genomics
Dovetail Genomics
Oxford Nanopore Technologies

TOWARDS A COMPLETE REFERENCE OF HUMAN GENOME DIVERSITY

Human Pangenome